DOCUMENT RESUME

ED 212 656 TM 820 048

AUTHOR Baker, Eva; Polin, Linda
TITLE Ongoing Studies in Domain-Referenced Content Validity: First Look at the "Judgment" Issue.
INSTITUTION California Univ., Los Angeles. Center for the Study of Evaluation.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE Jun 78
NOTE 16p.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Questionnaires; *Rating Scales; *Test Construction; Test Format; *Test Items; *Test Validity
IDENTIFIERS Fuzzy Set Theory; *Item Rating Sheet

ABSTRACT

The validity studies planned for the Test Design activities deal primarily with the appropriateness of items generated for a domain. Previous exploratory work in the field related to overall test content appropriateness ratings has not been satisfactory. Studies which are solely based on correlational data suffer from confounding with instructional experience. For this reason, the research uses a validity procedure providing refined estimates of congruency using theoretically based judgmental methods and empirical information. The document develops an instrument to rate features of individual test items, stemming from concerns with descriptive validity and addressing the quality of the match between an item on a test and the overall test domain. The idea of rating the degree of this match was developed through "fuzzy set" theory in which it is possible to specify strength of membership, or degree of inclusion, for a set member along a range of decimal values from 0 to 1. (Author/CE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED212656

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
[X] This document has been reproduced as received from the person or organization originating it.
[ ] Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

DELIVERABLE: June 1978

STUDIES IN MEASUREMENT & METHODOLOGY

WORK UNIT 1

Eva Baker
Project Director

ONGOING STUDIES IN DOMAIN-REFERENCED CONTENT VALIDITY: FIRST LOOK AT THE "JUDGMENT" ISSUE

Eva Baker
Linda Polin

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

G.S. Gray

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM820048

CENTER FOR THE STUDY OF EVALUATION
Graduate School of Education
University of California - Los Angeles

"The activity which is the subject of this report was supported in part or in whole by the National Institute of Education, Department of HEW. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred."

The validity studies planned for the Test Design activities have continued to address the issue of the appropriateness of items generated for a domain. Previous exploratory work in the field related to overall test content appropriateness ratings has not been satisfactory. Studies which are solely based on correlational data suffer from confounding with instructional experience. Thus, we have attempted to use a validity procedure that provides refined estimates of congruency using both theoretically based judgmental methods and empirical information.

Our development of an instrument to rate features of individual test items arises from concerns with descriptive validity and addresses the quality of the match between an item on a test and the overall test domain. The idea of rating the degree of this match was developed through "fuzzy set" theory in which it is possible to specify strength of membership, or degree of inclusion, for a set member along a range of decimal values from 0 to 1.

Procedure

This notion of degree of belongingness and the accompanying fuzzy set rating scale was applied to rating the match between each of a series of available test items and a domain specification. A group of six raters experienced in test evaluation and construction were gathered together to give their judgments of the match between each item in a series of items and a related domain specification. There were three domain specifications covering topics in reading (main idea, literal detail) and in mathematics (monetary values). There were ten items to be rated for each specification. Raters were given the following directions:

"Read the domain specification. Then look at the first test item. Consider how strongly you feel it belongs within the set of items described by the domain specification and indicate your judgment of "belongingness" by betting some proportion of 100 points or "chips." If you feel an item belongs completely, it should be assigned 100 points. An item you feel does not belong at all should receive 0 points. After you have assigned an amount for an item, go on to the next. Please make your judgment of each item independently of your judgments for the others. Please note that elements of the domain specifications (e.g., the sample item) have been deliberately ommitted."

Although raters had no problem making the fuzzy set judgments, there was no consistency within or among raters in the item and domain attributes upon which judgments were made. It became evident that there was a need for a systematic and uniform method of judging item qualities. In drafting the Item Rating Sheet (IRS), we felt we could identify component dimensions of item and subtest properties and provide an instrument encompassing these dimensions in a rating form conducive to making the difficult judgments described above.

Description Sheet

The IRS begins to address this judgment problem by distinguishing among item dimensions. At present the major dimensions we are studying include:

- Subject Matter tapped by the item, in terms of the discreteness of content and its instructional relevance.

- Cognitive Processing Complexity refers to the type and level of sophistication of thinking required to correctly answer the item.

- Syntactic/Semantic Level of Difficulty concentrates primarily upon the structural and vocabulary qualities affecting item difficulty.

- Format is concerned with level of difficulty based upon the method of presentation and includes such qualities as directions, set-up, graphs or other illustrations and symbols.

The following pages contain the present draft of the IRS instrument. The semantic/syntactic section is not yet in draft form and has been ommitted. Allowing for external review and revision processes, we anticipate completing these sections of the IRS by November 1, 1978.

Following final revision, raters will be trained to use the IRS instrument to make item and subscale judgments. These trained raters will participate in a series of studies on item clustering and degree of match between items and domain specifications. The problems and successes of this application of the "fuzzy set" concept will lead then to the integration of this judgmental approach to item/subscale quality with empirical procedures. In a study looking at the relationship between actual student performance on test items and our judgments of the descriptive validity of those items, we hope to substantiate this alternative conception of descriptive validity at the item level.

Application

We expect this procedure to be of value to school personnel charged with selecting or constructing tests for local proficiency tests.

Test Name_____

Page and Item Number_____

Rater Name_____

Date_____

I. ITEM SUBJECT MATTER

A. Topic

1. What is your perception of the broad subject matter area of this item?

a. Mathematics options operations_____

b. Reading_____

c. Spelling_____

d. Grammar_____

e. Mathematics Word Problems and/or Graphs

_____

f. Science_____

g. History/Government_____

h. Other Social Sciences_____

i. Map Reading_____

2. Within that area, what aspect/dimension/sub-heading of the general subject is being evaluated?_____

_____

3. What, if any, skills outside the subject area are also necessarily being tested for? (e.g., reading comprehension in mathematics word problems, map reading skills)

_____

_____

_____

B. Parameters

1. Based upon the item, what do you think is the general objective the item measures?_____

_____

_____

_____

2. (For Teachers Only): Based upon classroom objective(s) of yours, for what purpose would you be able to use this item?

_____

_____

_____

_____

II. Level of Cognitive Complexity

A. Operations to correctly interpret stimulus (question or item stem)

Y/N 1. Must the student rearrange information by a particular characteristic? If yes, check which operation(s) apply:

___ isolate relevant from irrelevant information

___ breakdown into components, subsets or categories

___ group discrete components together

Y/N 2. Must the student reinterpret information in stimulus (question or item stem)? If yes, check which operation(s) apply:

___ use context to derive meaning

___ change stimulus to new but equivalent form

B. Operations to make the correct response to the stimulus/question or item stem:

Y/N 1. Is the information in stimulus directly useable, explicit? (If yes, skip questions 2 and 3)

Y/N 2. Must the student identify and match:

___ relationships

___ patterns

___ meanings

Y/N 3. Must the student deduce a main idea or a "rule" governing a pattern (e.g., inference, number series):

___ use main idea to select

___ use "rule" to predict next case

___ use "rule" to generate case

Y/N 4. Must the student perform a known and directly applicable operation (e.g., multiplication, plural suffix)?

Y/N 5. In selection responses must the student select the one correct response?

C. Overall item and response complexity

1. What modes of information are presented in the item? (You may check more than one).

___ figural (sympolic or pictoral)

___ numeric

___ verbal

2. Is the solving operation specified in the task instructions (directions to test taker)?

___ yes, explicitly

___ implied but not explicitly stated

___ not specified or implied in instructions

3. Rate the importance of each of the following information sources for answering the item. (Use: 1 =very important; 4=not important)

___ exposure to formal schooling

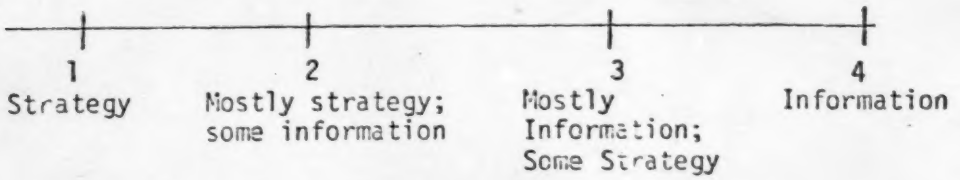
___ specific instruction or curriculum

___ social, cultural or familial experience

___ inate reasoning ability

___ verbal fluency

4. Rate the importance of strategy employed and/or the information recalled in aiming at the correct response.



5. As far as you can tell, is there only one way (conceptually) to answer the item correctly?

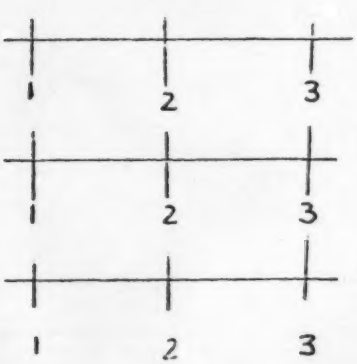
   a. yes, one way only

   b. no, more than one way but all solution/response strategies reflect the same skill and level of difficulty

   c. no, more than one way of differing skill type and/or level of difficulty

   d. no, there are no right or wrong answers

6. Overall, how would you rate the clarity of the item on:



Ambiguity - presence of 2 or more plausible meanings

Vagueness - absence of any definite inherent meaning

Complexity - presence of a meaning that is so complex or compounded that the reader has to work hard to put it all together

III. FORMAT

A. Directional Clarity

1. The instructions are given in (check all that apply):

___ written verbal form

___ pictorial form

___ oral form

2. May the test taker ask questions?

___ yes

___ no

___ unsure

3. How would you rate the general syntactic character of the directions:

simple sentences
(subject-verb, direct objectives, modifiers)

complex sentences
(compound sentences dependent clauses, antecedents.)

4. Is an illustrative and representative sample item given?

___ yes (effective, clarifying, similar format)

___ no (not present or not illustrative)

5. Do the instructions include information on how to indicate the selected response?

___ yes

___ no

6. Is the item independent of other test items such that a wrong response on preceeding or successive items will not have a positive or negative biasing effect?

___ yes

___ no

7. Does the semantic difficulty of the instructions section correspond with that of the items to which it refers?

___ yes

___ no

B. Item Clarity

1. Do the instructions include a graph, map, symbol key, or any other information which must be used in correctly responding to the item?

___ yes, If yes, is this information presented along side the item, or clearly/visibly repeated on page(s) following the initial presentation? Y/N

___ no

2. Is it clear which items belong with which given information (instructions, and where appropriate, additional information)?

___ yes

___ no

3. What type of physical response is required?

a) ___ selected from alternatives

b) ___ constructed in space provided

c) ___ one response only

___ more than one response per item

d) ___ oral

___ written

___ both oral and written

4. What is the complexity of the response format (circle, underline)?

___ simple indication or mark

___ simple constructed response (e.g., word or number(s))

___ complex constructed response, verbal (e.g., paragraph or sentence(s))

___ complex constructed response, nonverbal (e.g., drawing, diagram, equation)

Test Name_____

Page and Item Number_____

Rater Name_____

Date_____

I. ITEM SUBJECT MATTER

A. Topic

1. What is your perception of the broad subject matter area of this item?

a. Mathematics options operations_____

b. Reading_____

c. Spelling_____

d. Grammar_____

e. Mathematics Word Problems and/or Graphs

_____

f. Science_____

g. History/Government_____

h. Other Social Sciences_____

i. Map Reading_____

2. Within that area, what aspect/dimension/sub-heading of the general subject is being evaluated?_____

_____

3. What, if any, skills outside the subject area are also necessarily being tested for? (e.g., reading comprehension in mathematics word problems, map reading skills)

_____

_____

_____

B. Parameters

1. Based upon the item, what do you think is the general objective the item measures?_____

_____

_____

_____

2. (For Teachers Only): Based upon classroom objective(s) of yours, for what purpose would you be able to use this item?

_____

_____

_____

_____